Parental Somatic Mosaicism Is Underrecognized and Influences Recurrence Risk of Genomic Disorders

Ian M. Campbell,[1,11] Bo Yuan,[1,11] Caroline Robberecht,[2] Rolph Pfundt,[3] Przemyslaw Szafranski,[1] Meriel E. McEntagart,[4] Sandesh C.S. Nagamani,[1,5] Ayelet Erez,[1,5] Magdalena Bartnik,[6] Barbara Wiśniowiecka-Kowalnik,[6] Katie S. Plunkett,[1] Amber N. Pursley,[1] Sung-Hae L. Kang,[1] Weimin Bi,[1] Seema R. Lalani,[1,5] Carlos A. Bacino,[1,5] Mala Vast,[4] Karen Marks,[4] Michael Patton,[4] Peter Olofsson,[7] Ankita Patel,[1] Joris A. Veltman,[3] Sau Wai Cheung,[1] Chad A. Shaw,[1] Lisenka E.L.M. Vissers,[3] Joris R. Vermeesch,[2] James R. Lupski,[1,5,8,9,*] and Paweł Stankiewicz[1,10,*]

New human mutations are thought to originate in germ cells, thus making a recurrence of the same mutation in a sibling exceedingly rare. However, increasing sensitivity of genomic technologies has anecdotally revealed mosaicism for mutations in somatic tissues of apparently healthy parents. Such somatically mosaic parents might also have germline mosaicism that can potentially cause unexpected intergenerational recurrences. Here, we show that somatic mosaicism for transmitted mutations among parents of children with simplex genetic disease is more common than currently appreciated. Using the sensitivity of individual-specific breakpoint PCR, we prospectively screened 100 families with children affected by genomic disorders due to rare deletion copy-number variants (CNVs) determined to be de novo by clinical analysis of parental DNA. Surprisingly, we identified four cases of low-level somatic mosaicism for the transmitted CNV in DNA isolated from parental blood. Integrated probabilistic modeling of gametogenesis developed in response to our observations predicts that mutations in parental blood increase recurrence risk substantially more than parental mutations confined to the germline. Moreover, despite the fact that maternally transmitted mutations are the minority of alleles, our model suggests that sexual dimorphisms in gametogenesis result in a greater proportion of somatically mosaic transmitting mothers who are thus at increased risk of recurrence. Therefore, somatic mosaicism together with sexual differences in gametogenesis might explain a considerable fraction of unexpected recurrences of X-linked recessive disease. Overall, our results underscore an important role for somatic mosaicism and mitotic replicative mutational mechanisms in transmission genetics.

Introduction

New DNA mutations are understood to occur between generations, for example, when normal parents bear a child with a dominant disorder. Such a mutation could occur during a germline meiotic cell division, resulting in the sperm or egg; however, many mutations arise not in germ cells but rather during the ~10^{16} mitotic cell divisions required to generate an adult organism of ~10^{14} cells.[1] Owing to the inherent instability of the human genome, more than one mutation is generated per mitotic division.[2,3] The outcome of mitotic errors is mosaicism, defined as the presence of different cell populations with distinct genotypes within one individual.[4–8] Germline mosaicism in one of two healthy parents has long been invoked to explain recurrence of rare dominant disorders, even before the advent of molecular techniques.[2,3] The increasing sensitivity and resolution of genomic technologies has enabled the identification of mosaicism for both single-nucleotide variants (SNVs) and copy-number variants (CNVs) and has elucidated the contribution of mosaic mutations to a number of human genetic diseases.[4–8] Interestingly, in some instances, such as Proteus syndrome (MIM 176920), pathogenic alleles are found exclusively in the mosaic state; constitutional mutations are presumably embryonically lethal.[9] Moreover, combined somatic and germline mosaicism has been identified in parents of individuals with a number of genetic conditions,[4] thus raising the possibility that mosaic individuals might be detected by routine blood tests rather than direct examination of germ cells. Recent emerging evidence also suggests that somatic mosaicism occurs in apparently healthy individuals and increases with age.[10–12] Yet, despite its considerable impact on human health, systematic population-level studies of mosaicism contributing to the transmission of genetic disease are lacking.

An increasingly recognized class of disease-associated mutations is CNVs, generally classified as recurrent or nonrecurrent genomic rearrangement events that cause DNA-dosage changes or deviations from the normal diploid state. Nonrecurrent CNVs have breakpoints that are not clustered in particular genomic regions and are

[1]Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [2]Centre for Human Genetics, University Hospital, Katholieke Universiteit Leuven, Leuven 3000, Belgium; [3]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Medical Center, Nijmegen 6500 HB, the Netherlands; [4]Department of Medical Genetics, St. George's University of London, Cranmer Terrace SW17 0RE, UK; [5]Texas Children's Hospital, Houston, TX 77030, USA; [6]Department of Medical Genetics, Institute of Mother and Child, Warsaw 01-211, Poland; [7]Mathematics Department, Trinity University, San Antonio, TX 78212, USA; [8]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [9]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [10]Institute of Mother and Child, Warsaw 01-211, Poland
[11]These authors contributed equally to this work
*Correspondence: jlupski@bcm.edu (J.R.L.), pawels@bcm.edu (P.S.)
http://dx.doi.org/10.1016/j.ajhg.2014.07.003.

usually formed by microhomology-mediated DNA-replication errors (fork stalling and template switching or microhomology-mediated break-induced replication) during cell divisions or by nonhomologous-end joining, whereas recurrent CNVs refer to genomic-disorder-associated CNVs with clustered breakpoints, usually located in flanking low-copy-repeat regions and mediated by nonallelic homologous recombination.[13] Directed molecular studies have detected somatic mosaicism and implied obligate carrier gonadal mosaicism in parents of individuals with apparently de novo CNVs by using karyotype, fluorescence in situ hybridization (FISH), Southern blot, or PCR analyses of genetic material derived from peripheral-blood cells.[14–17] We previously identified very low-level somatic mosaicism and inferred germline mosaicism for CNVs in parents of individuals with genomic disorders,[18,19] wherein the pathogenic mutant alleles were not detected by standard clinical assays but were instead identified by the increased sensitivity of CNV-specific PCR. Thus, we hypothesized that somatic mosaicism for CNVs that also contribute to the germline might be more common than currently appreciated.

Subjects and Methods

Subjects

Human subjects research was approved by the institutional review board of Baylor College of Medicine (BCM) under protocol H28088. Informed consent was obtained from each research participant. We queried our clinical databases for families of individuals with genomic deletions that were determined to be de novo. The deletions in this study were originally ascertained by a variety of clinical diagnostic SNP and array comparative genomic hybridization (aCGH) platforms. Parents of individuals tested at the BCM Medical Genetics Laboratory, at Katholieke Universiteit Leuven University Hospital, and at the Institute of Mother and Child had peripheral-blood lymphocytes interrogated by FISH or aCGH. Parents of individuals tested at Radboud University Medical Center had peripheral-blood genomic DNA interrogated by 250K SNP microarray. In each case, no evidence of the child's deletion could be found in either parent. Genomic DNA isolated from peripheral blood (or from saliva in two parental pairs) from each affected individual and his or her parents was then subjected to further testing. This study focused on deletion CNVs interpreted to be pathogenic or potentially pathogenic by the respective diagnostic laboratories.

aCGH

Custom region-specific high-resolution CGH arrays were designed on the basis of information gained from each affected individual's clinical microarray studies with the use of the web-based eArray software (Agilent Technologies). Genomic intervals to be interrogated for fine mapping of breakpoint junctions were estimated from clinical arrays and the transitions from normal copy number to loss of copy number, which signified CNV deletion. For array design, a 200,000 bp window was placed around the uncertainty region of each breakpoint (Figure S1A, available online). Probes were then placed automatically within each interval for a final average spacing of approximately one probe per 500 bp. Probes for multiple affected individuals were tested on each 8-plex 60k feature CGH microarray (Agilent Technologies). Digestion, labeling, and hybridization of genomic DNA were performed according to the manufacturer's instructions. Because only two individuals had overlapping deletions, chromosomally sex-matched individuals were hybridized to each other. Data were analyzed with Agilent Genomic Workbench Software.

Long-Range PCR

Long-range PCR (LR-PCR) primers were designed for the personal genome of each affected individual on the basis of custom aCGH data such that primers were located within the 500 bp immediately flanking the deletion (Figure S1B). If repetitive elements prevented the design of primers in these regions, more distant primers were utilized. LR-PCR was employed to amplify junction fragments of deletions from affected individuals' peripheral-blood genomic DNA according to the manufacturer's protocol (Takara Bio). Sanger sequencing of the amplicons containing breakpoint junctions confirmed their localization to the genomic region where the deletion CNV mapped (Lone Star Labs). CNV coordinates have been deposited in the Database of Genomic Variants Archive and are available under the accession number estd211. Once a specific breakpoint fragment was identified, 1 μg of genomic DNA from each of the affected individuals' parents was subjected to the same LR-PCR reaction. To ensure completion of the reaction, temperature cycling was completed 45 times. LR-PCR reactions were then visualized by electrophoresis in a 1% agarose gel with ethidium bromide. Given that the mass of the diploid female human genome is ~6.4 pg[20] and given the assumption of 50% PCR efficiency, a level of mosaicism of ~1 in 75,000 could potentially be detected.

Multiplex LR-PCR

In the case of family 2, multiplex LR-PCR was utilized for the determination of the breakpoint. LR-PCR primers were designed such that one primer was synthesized for each 7 kb of uncertainty at the affected individual's breakpoints. A total of 19 primers were added to the 25 μl LR-PCR reaction at a final concentration of 500 nM each. The remaining reaction conditions were unchanged from the manufacturer's (Takara Bio) protocol. Sanger sequencing was then used for identifying the primer pair resulting in the specific fragment. Separate mixtures of all forward and reverse primers were provided for sequencing (Lone Star Labs).

Droplet PCR

In families with a mosaic parent, peripheral-blood DNA from available individuals was subjected to droplet PCR specific to both the familial deletion and a normal diploid segment of the genome according to the manufacturer's recommendation (QX200 Droplet Digital PCR System, Bio-Rad Life Science Research). Data were analyzed with QuantaSoft v.1.4 software (BioRad Life Science Research). Mosaic individuals were normalized to the number of normal diploid genome copies identified in the sample; we also controlled for PCR efficiency by assuming that 100% of the affected offspring's cells harbored a mutation.

Cell-Division Modeling

We used a two-type Galton-Watson model for mutations in a clonal lineage where mutations were assumed to have no effect on cell fitness and to occur at a constant rate per mitotic division.[21] We assumed a total of 30 generations of division in the

female line and 400 generations in the male line.[22,23] We modeled an exponential stochastic expansion of the cell population to allow for both cell death and division in generations 1–30 (Figure S2). For modeling the male germline, we defined a critical process where self-renewal, division, and death were in equilibrium to model divisions in spermatogenesis for generations 30–400.[23] For all analyses, we chose a constant fitness ($p = q = r = s = 0.9$).[21] We used the sampling formulas from Olofsson and Shaw[21] together with our analytical results regarding the probability generating functions to compute the expected mean and variance of the proportion of mutant gametes in the female germline. To create a more comprehensive set of results that consider the switch from clonal expansion to stable self-renewal in the male germline and to capture the change in recurrence risk conditional on observing affected offspring, we extended the past work by using a large-scale Monte Carlo approach.

We performed over 1×10^9 forward simulations representing the process of gametogenesis implemented in the R Statistical Programing Language (R Core Development Team). During each generation, we considered stochastic clonal division or self-renewal coupled with random mutation in accordance with the probabilities listed in Figure S2. To enable later calculation of recurrence risk and the change in recurrence risk dependent on observed transmission of a mutant allele and the observation of parental mosaicism, we also recorded the mitotic division, in which each mutation present in the ultimate gamete pool was generated. The model utilized simple independent Bernoulli trials for each division event, which accumulated to binomial random variables and multinomial variables when we tracked the cumulative offspring of normal cells and mutants arising in each successive generation.

We performed preliminary analysis on mutation rates of 1×10^{-8}, 1×10^{-10}, and 1×10^{-12} mutations per mitotic event because the mutation rates of randomly selected specific CNVs are poorly understood. Although mutation rate strongly influences the risk of having a first affected offspring in the general population, we identified a minimal influence of mutation rate on recurrence risk of parents with affected offspring. Therefore, to facilitate analysis on large numbers of realizations with mutants, we subsequently used a constant rate of 1×10^{-8} per mitotic event. Using this mutation rate, we were able to generate over 2×10^8 realizations of gametogenesis with at least one mutant present in the final pool of gametes. Because we were focused on events in which parents were unaffected and were prescreened for high-level mosaicism, our analyses excluded the possibility of mutation at extremely early epochs in development (generations 1–4). Our modeling of meiosis resulted in exactly 50% of gametes affected for each mutant diploid germ cell.

Analysis of Recurrence Risk

The methodology we developed in a previous study[21] allows for the exact determination of the expected frequency of mutants in a clonal population initiated from a single normal cell. This expected frequency is unconditional on the observation of a mutation transmitted to affected offspring or the knowledge of parental mosaicism or its absence. To update the expected proportion of mutants (the recurrence risk) conditional on sampling a mutant gamete, $E[\theta \mid T > 0]$, we used our Monte Carlo results and examined the joint distribution of sampling a mutant when it originated at generation k. We reasoned that we could compute the expected proportion of mutants conditional on sampling by conditioning on the generation of origin of a sampled mutation and then summing over the generation of origin and multiplying by the empirical Monte Carlo probability of sampling a mutant that arose in generation k. We observed that the distribution of origin of a sampled mutation was uniform across k for the female line (Figure S3A) and biphasically uniform in the clonal expansion and stable self-renewal phase for males. Therefore, we used a uniform distribution for the generation of origin in our calculations and the following formula:

$$
\begin{aligned}
E[\theta_{\text{mom}} \mid T_{\text{mom}} > 0] &= E[E[\theta_{\text{mom}} \mid T_{\text{mom}} > 0,\ O = k]] \\
&= \sum_{k=i}^{j} E[\theta_{\text{mom}} \mid T_{\text{mom}} > 0,\ O = k] \times P[O = k] \\
&\approx \sum_{k=i}^{j} \left[\sum_{\text{sim}\ X_k > 0} \left[\frac{\sum \boldsymbol{X}}{Y + \sum \boldsymbol{X}} \times \frac{X_k}{Y + \sum \boldsymbol{X}} \right] \times \frac{1}{j - i + 1} \right],
\end{aligned}
$$

where θ_{mom} is the expected proportion of mutant gametes in the mother, T_{mom} is the type ($T = 0$ for normal, $T = 1$ for mutant) of the mother's sampled gamete, O is the mitosis of origin of the mutation, P is the probability that a sampled gamete originated in generation k, X_k is the number of mutant gametes arising in generation k, $\boldsymbol{X}$ is a vector of the number of mutants originating in each generation, and Y is the number of normal gametes. Using this formula, we were able to determine the expectation of the proportion of mutants given that a sampled mutant arose in each generation k (Figure S3B). Importantly, the sampled mutation tends to arise in processes where the mutant subpopulations are larger in relation to typical clonal expansion initiated from each generation k, a phenomenon known as size-biased sampling.[24] We developed a similar equation for spermatogenesis. Our simulation results then allowed us to determine the overall expectation of mutant proportion—which is the chance that a second mutant gamete is sampled when these factors are taken into account.

Parent of Origin

To determine the change in recurrence risk, we considered the ratio of mutant proportion unconditionally to that determined by observing an affected child. If θ_p is the proportion of mutants in the parent of origin, the recurrence risk can be thought of as the conditional expectation of θ_p given the observation of a mutant gamete. In the absence of information on the parent of origin, we must integrate (sum) over the two possible parents. We also conditioned on the generation of origin of the transmitted mutation during development of the parental germline. The formula for our approach is

$$
\begin{aligned}
E\left[\theta_p \mid T_{\text{mom}} + T_{\text{dad}} > 0\right] \approx\ & E[\theta_{\text{mom}} \mid T_{\text{mom}} > 0] \times P[T_{\text{mom}} > 0] \\
& + E[\theta_{\text{dad}} \mid T_{\text{dad}} > 0] \times P[T_{\text{dad}} > 0],
\end{aligned}
$$

where $P[T_{\text{mom}} > 0]$ and $P[T_{\text{dad}} > 0]$ are the unconditional probabilities of sampling a mutant gamete from the mother and father, respectively. We computed the expectations on the right-hand side of the equation as described above. The simulations reveal that information contributed by sampling a mutant is asymmetric in mothers and fathers because of the different processes that underlie gametogenesis in the sexes.

Literature Review

We searched PubMed to identify cases of familial recurrence of apparently de novo CNVs by searching for "CNV familial

recurrence," "deletion familial recurrence," "duplication familial recurrence," and "sibling mosaicism." We reviewed each report to verify that (1) the mutation was a "nonrecurrent" CNV, (2) the authors used a technique that could exclude a balanced translocation, and (3) the authors used a technique that could unambiguously identify the parent of origin. Although there is no evidence that this represents a biased sample, we cannot fully exclude an ascertainment bias due to the generally lower availability of fathers in genetic testing.

Results

As a proof of principle, we tested our hypothesis experimentally by investigating two families, each of whom had an unexpected recurrence of a genomic disorder. Family 1 was referred for assessment of three children with suspected Smith-Magenis syndrome (SMS [MIM 182290]) and a common mother but two different fathers (Figures 1A and 1B). aCGH validated for clinical diagnosis revealed in all three children a causative ~40 kb deletion encompassing the final two exons of *RAI1* in chromosomal region 17p11.2 but did not identify a carrier parent. Utilizing the increased resolution of a custom-designed region-specific high-resolution aCGH (Figure 1A) and an individual-specific LR-PCR assay (Figure 1B), we amplified and sequenced the breakpoint junction of the deletion CNV (Table 1; Figure S4), revealing a complex genomic rearrangement most likely due to a DNA-replication error. LR-PCR amplification of the mother's peripheral-blood DNA resulted in an identically sized band of lower intensity, consistent with somatic mosaicism (Figure 1B). Droplet PCR analysis of familial DNA samples normalized to the affected individuals revealed mutant alleles in 25.1% of maternal blood cells (Figure 1C), underscoring the limited ability of currently available clinical tests to detect mosaicism. Likewise, individual-specific LR-PCR analysis of family 2—in which two brothers have developmental delay, microcephaly, and dysmorphic features (Figure S5A) caused by identical 1q43–q44 deletions containing the haploinsufficient *AKT3* (MIM 603387)—again revealed low-level mosaicism (in this case, 3.4% of blood cells tested) in an apparently noncarrier father previously tested by clinically validated FISH (Table 1; Figures S4 and S5).

Given the technical capability to detect somatic mosaicism in families with unexpected recurrences of dominant disease, we hypothesized that mosaicism might also be present in parents with a child affected by a simplex genomic disorder. We prospectively tested peripheral-blood DNA from 100 parental pairs for the nonrecurrent, interstitial, and pathogenic or potentially pathogenic deletion CNV detected in their offspring (Table S1; Figure S6). In each case, the rearrangement had been previously categorized as de novo by clinical testing using aCGH, SNP arrays, or FISH analyses. For each family, we generated deletion-specific LR-PCR primers by prior narrowing of the genomic coordinates of the breakpoints by using high-resolution custom-designed aCGH. Sequence analysis of 91 deletion CNV breakpoint junctions in the affected individuals revealed potential hallmarks of replicative mechanisms[13] at a number of the breakpoint junctions (Figure S7; Tables S2 and S3).

Subsequent parental testing identified two mothers and two fathers mosaic for their child's deletion (Figure 2). The deletion size, gene content, and breakpoint mutational signature for each family, as well as available mosaic fractions measured from parental-blood DNA, are presented in Table 1 and Figures S4 and S8. We also incidentally identified a balanced insertional translocation (Figure S9).[25] Contamination of the parental DNA with that of the affected individuals is unlikely. Each result was independently confirmed by the diagnostic laboratory providing the sample. Moreover, in family 3, which had an affected son with a mosaic mother, we did not detect *SRY*-specific amplicons from the mother's peripheral-blood DNA (data not shown). The deletion breakpoints in family 6 were located in directly oriented *Alu* repetitive elements. The breakpoint amplicon was detected in the father, but not the mother, independently by two laboratories; nevertheless, a repetitive-element-mediated PCR artifact cannot be completely excluded. Overall, the most parsimonious explanation for the findings in families 3–6 is that the parents are combined somatic and germline mosaics and that they are at a higher risk for recurrence, as observed retrospectively in families 1 and 2.

Discussion

Human primordial germ cells (PGCs) are derived from pluripotent epiblast cells and are segregated in the dorsal yolk sac endoderm by day 24 of embryogenesis (Figure 3A).[26,27] Postzygotic mutations occurring during mitotic events between the 1-cell embryo stage and differentiation of the PGCs could contribute to the embryonic endoderm and mesoderm, the latter of which gives rise to hematopoietic stem cells (HSCs). These HSCs would in turn differentiate into nucleated blood cells, which we interrogated by our CNV-breakpoint-junction-specific LR-PCR assay. Prospective screening revealed low-level mosaicism in DNA obtained from parental blood in 4% of affected individuals (95% confidence interval = 1.1%–10.2%), suggesting that a considerable number of mutations arise during mitotic cell divisions in the previous generation and can be transmitted to offspring. This 4% rate is potentially an underestimate of the true rate of mosaicism, given that the families in our study were prescreened for the exclusion of clinically detectable mosaicism by existing methods and that some of the prospectively ascertained parents might be mosaic but have remained undetected because the mosaicism is present in cell lineages other than those tested here.

To further contextualize our observations, we developed a probabilistic model considering mutation during embryogenesis and gamete formation. Our framework models each mitotic and meiotic cell division between generations; the

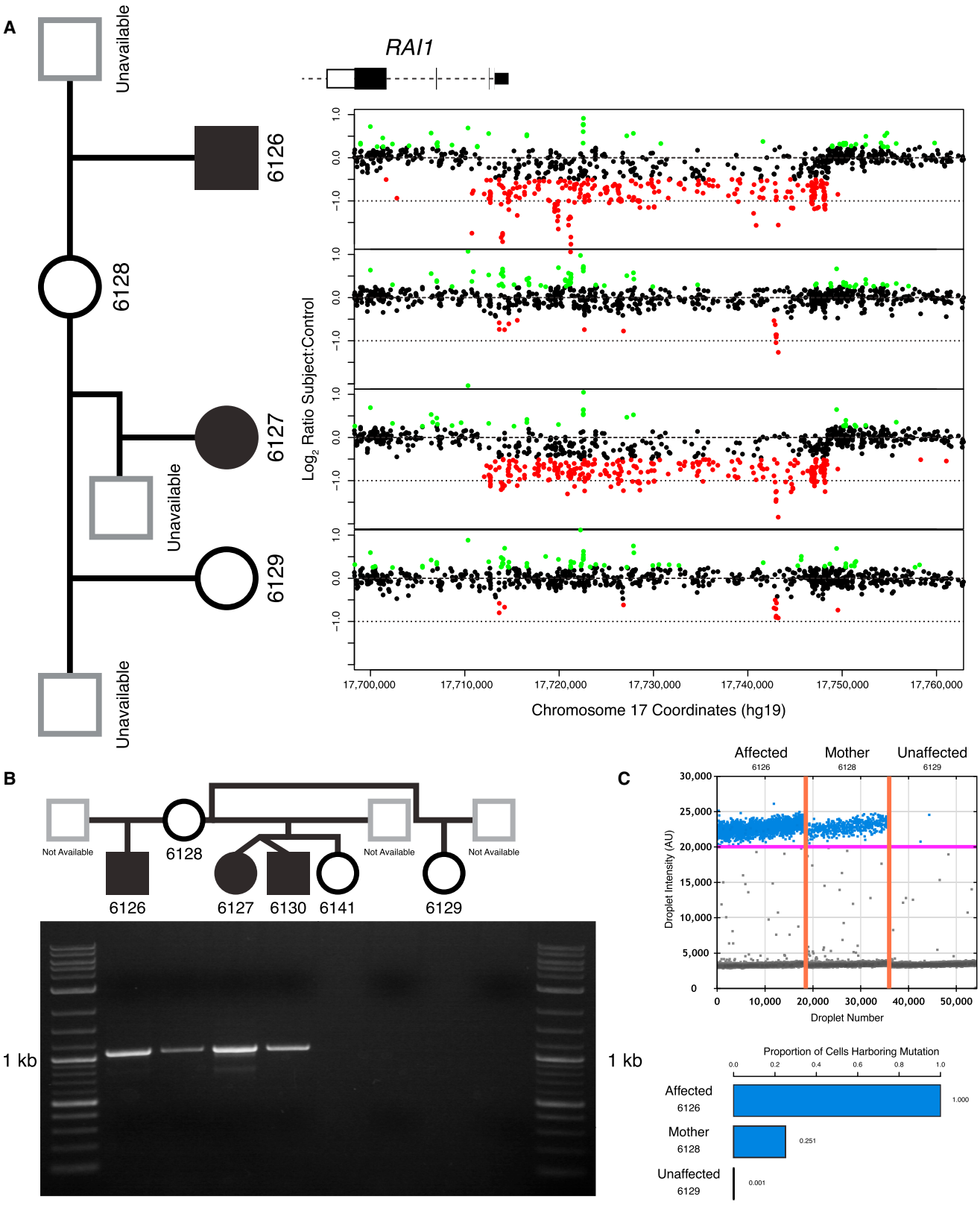

Figure 1. Low-Level Combined Germline and Somatic Mosaicism Inferred from Familial Recurrence of SMS

Family 1 was identified with three individuals suspected to have SMS and born to one mother but two different fathers.

(A) aCGH analysis of genomic DNA from the mother, two affected half siblings, and one unaffected half sibling. No detectable copy-number change was seen in the mother.

(B) LR-PCR analysis of genomic DNA from available family members. The familial deletion-specific amplicon segregated with the SMS phenotype in the children and was clearly visible from maternal peripheral-blood DNA.

(C) Digital PCR analysis of affected, maternal, and unaffected blood samples revealed mutations in 25.1% of maternal nucleated blood cells.

Table 1. Characteristics of Mosaic CNVs Identified in This Study

Family	Analysis	Coordinates	Size	Inheritance	Gene	MIM	Mutational Signature	Percent Mosaic
1	retrospective	chr17: 17,711,738–217,748,468	36.7 kb	maternal	*RAI1*	607642	39 bp normal intervening sequence, +1 bp identity, +2 bp microhomology	25.1%
2	retrospective	chr1: 242,263,612–244,559,673	2.3 Mb	paternal	*AKT3*[a]	611223	11 bp normal intervening sequence	3.4%
3	prospective	chr12: 23,585,878–23,829,423	244 kb	maternal	*SOX5*	604975	blunt breakpoint	9.0%
4	prospective	chr6: 75,502,925–75,867,029	364 kb	maternal	*COL12A1*	120320	2 bp microhomology	<1%
5	prospective	chr9: 119,474,386–119,587,581	113 kb	paternal	*ASTN2*	612856	3 bp microhomology	3.0%
6	prospective	chr2: 165,659,793–166,267,524	608 kb	paternal	*SCN2A*	182390	*Alu*/*Alu*, 181 bp 100% identity	<1%

All coordinates are according to the GRCh37/hg19 (2009) assembly.
[a]*AKT3* and eight other RefSeq genes.

number of divisions varies with sex and paternal age and is estimated to be ~30 for females and ~400 for a 30-year-old male (Figure 3B).[22,23,28] Although our model accommodates arbitrary changes in per-mitosis mutation rate, mutant cell fitness, and paternal age (see Material and Methods), we focused on mutagenic processes with constant rates in females and 30-year-old males for mutations with marginal cellular fitness effects. One prediction of our model is that somatically mosaic parents—that is, individuals who carry mutations that must have occurred during the first 15 mitotic events, before the segregation of the germline from the other cell lineages of the body (Figure 3B)—harbor on average 7 to 8 orders of magnitude more mutant gametes than the typical individual.

However, irrespective of mosaicism status, the observation of an affected child increases the expectation regarding the proportion of mutant gametes existent in the child's parents' germ cells because the parent is more likely to transmit a mutant gamete if he or she harbors a larger fraction of mutants than is typical for parents.[24] Conditioning on the observation of an affected child (see Material and Methods), our model estimates that average recurrence risk is ~0.1%, which is consistent with previous estimates.[29] In contrast, somatically mosaic parents of affected offspring are estimated to be, on average, at an approximately 512-fold (female parent of origin) to 3,312-fold (male parent of origin) higher risk than are parents in whom mutations are strictly confined to the germline. However, we caution that determining whether a parent harbors mutant cells only in the germline is not experimentally feasible at this time.

The equal number of somatically mosaic mothers and fathers in our prospective study is interesting (although not significantly different than the expected number, $p = 0.18$) given that only approximately 20% of apparently de novo nonrecurrent CNVs arise on the maternally inherited chromosome.[30] This previously observed sex bias in the parent of origin is most likely the result of additional mutations that arise during mitotic cell divisions of spermatogenesis,[31] and our modeling results are consistent with this hypothesis. One byproduct of this sexual dimorphism in gametogenesis is that mutations that arise in self-renewing spermatagonia are, on average, expected to expand into fewer gametes than are mutations that arise in cells during embryogenesis. Therefore, even though fewer mitotic mutations occur in mothers, those female individuals who do transmit a mutant gamete have, on average, a higher proportion of mutant gametes within their ovaries than the typical transmitting parent. Our model predicts that although mothers are the parent of origin in a minority of cases, they are at an order of magnitude higher risk of recurrence than typical transmitting fathers. This hypothesis is strengthened by a literature review of familial recurrence of apparently de novo CNVs in combination with our observations that identify the mother as the source of 17/21 (81%) rearrangements (Table S4). Thus, identifying the parent of origin of a mutation might have utility in estimating recurrence risk, even if the status of somatic mosaicism cannot be determined. For transmitted de novo mutations that cause X-linked recessive conditions, affected males necessarily inherit the new mutation from their mothers. Thus, the high rate of recurrence in families with apparently de novo mutations causing diseases such as Duchenne and Becker muscular dystrophies, hemophilia, and ornithine transcarbamylase deficiency[32] might be partially explained by somatic mosaicism.

Given that the commonly quoted risk of complications with invasive prenatal testing is ~0.5%, prenatal diagnosis is not routinely suggested for a second pregnancy following the birth of a child with a simplex case of a genomic disorder.[29,33] However, if screening for somatic mosaicism in parental blood is able to identify couples at substantially increased risk of a recurrence, prenatal diagnosis might be offered. Such prospective analyses might have changed recurrence-risk counseling for family 1, where strong evidence of mosaicism could have affected choice or management regarding additional pregnancies. In contrast to a number of methodologies routinely used in clinical diagnostics, our results suggest that using individual-specific breakpoint PCR to detect CNV mosaicism is more sensitive,

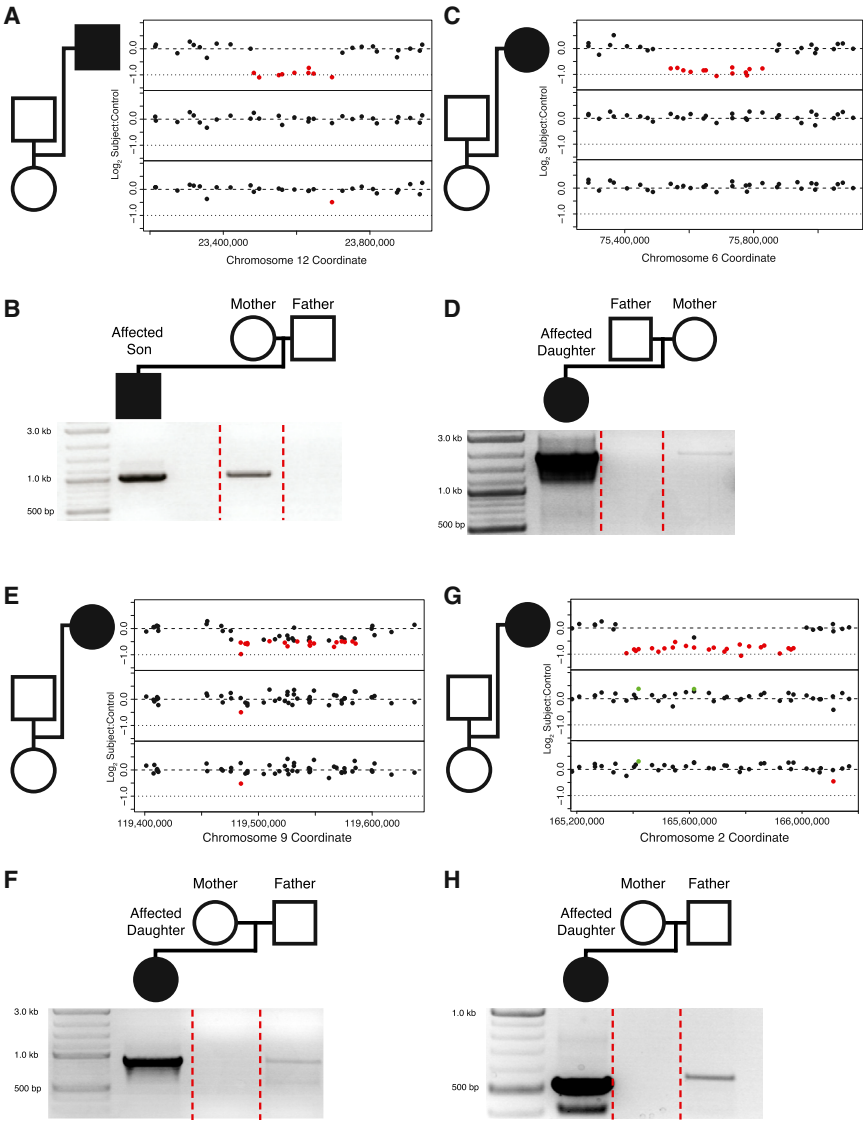

Figure 2. Low-Level Somatic Mosaicism Prospectively Identified in Four Families
(A) Microarray analysis of family 3 revealed an apparently de novo 250 kb deletion in chromosomal region 12p12.1.
(B) Deletion-specific LR-PCR in family 3 identified the amplicon detected in the affected son's peripheral-blood DNA also in the mother's DNA.
(C) Microarray analysis of family 4 revealed an apparently de novo 350 kb deletion in chromosomal region 6q13.
(D) Deletion-specific LR-PCR in family 4 identified the amplicon detected in the affected son's peripheral-blood DNA also in the mother's DNA.
(E) Microarray analysis of family 5 revealed an apparently de novo 100 kb deletion in chromosomal region 9q33.1.
(F) Deletion-specific LR-PCR in family 5 identified the amplicon detected in the affected daughter's peripheral-blood DNA also in the father's DNA.
(G) Microarray analysis of family 6 revealed an apparently de novo 608 kb deletion in chromosomal region 2q24.3.
(H) Deletion-specific LR-PCR in family 6 identified the amplicon detected in the affected daughter's peripheral-blood DNA also in the father's DNA.

is less expensive, and requires less infrastructure while also being less invasive than skin biopsy. We assessed the minimum detection limit of our LR-PCR assays by diluting DNA of an affected individual into that of a control subject and found that under ideal conditions, as few as one or two chromosomes bearing a deletion can be amplified to produce visible bands (Figure S10). One potentially useful downstream application for such sensitivity is breakpoint PCR to interrogate free fetal chromatin within maternal plasma to diagnose a recurrence noninvasively.

Our results also suggest that widely used tests for CNVs fail to detect a substantial fraction of low-level mosaicism. Similarly, many of the same tests most likely lack the precision to distinguish high-level somatic mosaicism (for example, 80%–100% of cells) from completely constitutional alternations. Thus, the variable expressivity and incomplete penetrance observed for some genetic conditions could be at least partially due to unrecognized mosaicism. Somatic mosaicism that arises during embryogenesis might also have an underappreciated contribution to cancer genetics. Somatic mutations could occur early during development and be harbored by multiple tissues, one of which could ultimately undergo malignant transformation. Such mutations could then be missed during analysis of other sources of genomic DNA.

In aggregate, our results suggest that a considerable number of apparently de novo mutations causing genomic disorders actually occur in the previous generation and can thus be recurrently transmitted to future offspring. Although our study assessed only CNVs, it is possible that any mutation occurring during mitosis can be transmitted in this manner. Higher genome resolution accompanying the shift of diagnostic testing toward massively parallel sequencing might allow rearrangement-specific LR-PCR to become an inexpensive yet sensitive test for CNV mosaicism. Likewise, sensitive and specific tests for SNVs are needed for identifying low-level mosaicism for other types of mutations. Such investigations could assist couples who are planning additional pregnancies after the birth of a child with a genomic disorder.

Supplemental Data

Supplemental Data include ten figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.07.003.

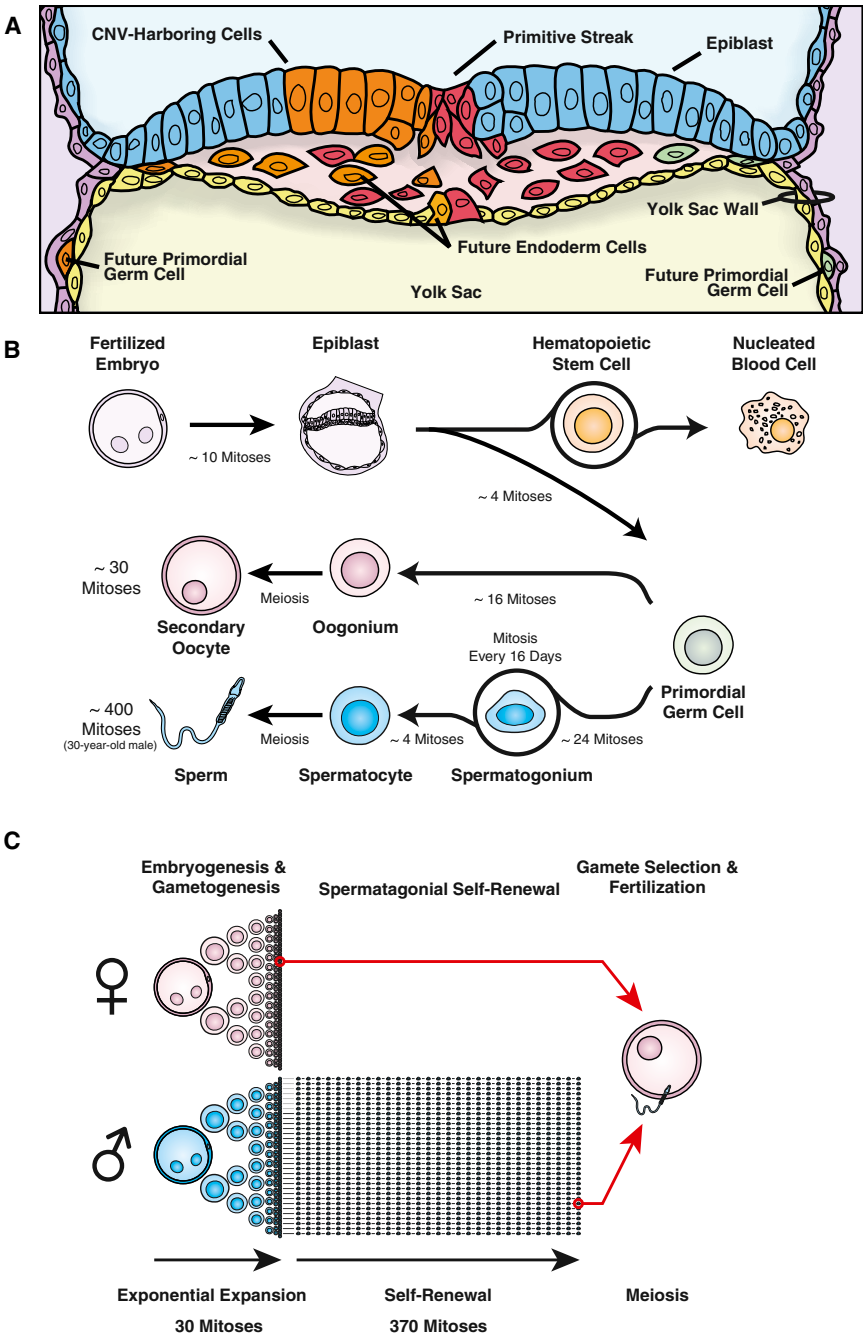

Figure 3. Human Germ Cell Development

(A) Epiblast cells invaginate during the third week of embryogenesis to form the future endoderm and mesoderm. Some dorsal endoderm cells near the allantois become situated in the wall of the yolk sac and later differentiate into primordial germ cells (PGCs). During the fourth and fifth weeks of gestation, these PGCs migrate to the primitive gonads to become gametes. If a CNV were to occur in an epiblast cell before the third week, later divisions could contribute to both PGC and mesoderm lineages, including hematopoietic stem cells (HSCs).

(B) Distribution of cell divisions during gametogenesis.

(C) Probabilistic model of development. Both males and females experience a stochastic exponential cell-expansion phase modeling embryogenesis and germ cell proliferation. In males, expansion is followed by a stochastic but nonexpanding process of self-renewal modeling spermatogenesis. A single sperm and egg are then randomly sampled after meiosis to fertilize an offspring. Mutations can arise in any cell division, contributing to the gamete pool, and are ultimately available to be transmitted to the next generation. Mutations that occur during the exponential-expansion phase can divide to comprise a larger proportion of the germ cell pool. In contrast, mutations that occur during the self-renewal phase expand into fewer mutant sperm because of asymmetric cell division.

Acknowledgments

We thank the families for their participation in this project. We thank Greet Peeters for laboratory support, John D. Belmont for insightful suggestions concerning X-linked disease, and La Donna Immken for assistance in subject recruitment. I.M.C. is a fellow of the Baylor College of Medicine (BCM) Medical Scientist Training Program (T32 GM007330-34) and was supported by a fellowship from the National Institute of Neurological Disorders and Stroke (F31 NS083159). S.C.S.N. is a recipient of the Clinical Scientist Development Award from the Doris Duke Charitable Foundation. This study was supported in part by grants from the Intellectual and Developmental Disabilities Research Center (P30 HD024064), the Baylor-Hopkins Center for Mendelian Genomics (U54HG006542), and the National Institute of Neurological Disorders and Stroke (R01 NS058529) to J.R.L. and by grants from the National Heart, Blood, and Lung Institute (R01 HL101975) and Polish Ministry of Science and Higher Education (R13-0005-04/2008) to P.S. J.R.L. holds stock ownership in 23andMe Inc. and Ion Torrent Systems Inc. and is a coinventor on multiple United States and European patents related to molecular diagnostics. The Department of Molecular and Human Genetics at BCM derives revenue from molecular genetic testing offered in the Medical Genetics Laboratories (http://www.bcm.edu/geneticlabs/).

Received: May 15, 2014
Accepted: July 8, 2014
Published: July 31, 2014

Web Resources

The URLs for data presented herein are as follows:

Database of Genomic Variants Archive (DGVa), http://www.ebi.ac.uk/dgva/

Genome Reference Consortium, http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

UCSC Genome Browser, http://genome.ucsc.edu

Accession Numbers

The DGVa accession number for the CNV coordinates reported in this paper is estd211.

References

1. Iourov, I.Y., Vorsanova, S.G., and Yurov, Y.B. (2010). Somatic genome variations in health and disease. Curr. Genomics *11*, 387–396.
2. Bakker, E., Van Broeckhoven, C., Bonten, E.J., van de Vooren, M.J., Veenema, H., Van Hul, W., Van Ommen, G.J., Vandenberghe, A., and Pearson, P.L. (1987). Germline mosaicism and Duchenne muscular dystrophy mutations. Nature *329*, 554–556.
3. Seshadri, R., Kutlaca, R.J., Trainor, K., Matthews, C., and Morley, A.A. (1987). Mutation rate of normal and malignant human lymphocytes. Cancer Res. *47*, 407–409.
4. Erickson, R.P. (2010). Somatic gene mutation and human disease other than cancer: an update. Mutat. Res. *705*, 96–106.
5. Lupski, J.R. (2013). Genetics. Genome mosaicism—one human, multiple genomes. Science *341*, 358–359.
6. Pham, J., Shaw, C., Pursley, A., Hixson, P., Sampath, S., Roney, E., Gambin, T., Kang, S.-H.L., Bi, W., Lalani, S., et al. (2014). Somatic mosaicism detected by exon-targeted, high-resolution aCGH in 10 362 consecutive cases. Eur. J. Hum. Genet. *22*, 969–978.
7. Forsberg, L.A., Absher, D., and Dumanski, J.P. (2013). Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. J. Med. Genet. *50*, 1–10.
8. Biesecker, L.G., and Spinner, N.B. (2013). A genomic view of mosaicism and human disease. Nat. Rev. Genet. *14*, 307–320.
9. Lindhurst, M.J., Sapp, J.C., Teer, J.K., Johnston, J.J., Finn, E.M., Peters, K., Turner, J., Cannons, J.L., Bick, D., Blakemore, L., et al. (2011). A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N. Engl. J. Med. *365*, 611–619.
10. Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C., et al. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. *44*, 642–650.
11. Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodríguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., Horner, M.-J., et al. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. Nat. Genet. *44*, 651–658.
12. Forsberg, L.A., Rasi, C., Razzaghian, H.R., Pakalapati, G., Waite, L., Thilbeault, K.S., Ronowicz, A., Wineinger, N.E., Tiwari, H.K., Boomsma, D., et al. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. Am. J. Hum. Genet. *90*, 217–228.
13. Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. Nat. Rev. Genet. *10*, 551–564.
14. Higuchi, M., Kochhan, L., and Olek, K. (1988). A somatic mosaic for haemophilia A detected at the DNA level. Mol. Biol. Med. *5*, 23–27.
15. Bakker, E., Veenema, H., Den Dunnen, J.T., van Broeckhoven, C., Grootscholten, P.M., Bonten, E.J., van Ommen, G.J., and Pearson, P.L. (1989). Germinal mosaicism increases the recurrence risk for 'new' Duchenne muscular dystrophy mutations. J. Med. Genet. *26*, 553–559.
16. Milewicz, D.M., Guo, D.-C., Tran-Fadulu, V., Lafont, A.L., Papke, C.L., Inamoto, S., Kwartler, C.S., and Pannu, H. (2008). Genetic basis of thoracic aortic aneurysms and dissections: focus on smooth muscle cell contractile dysfunction. Annu. Rev. Genomics Hum. Genet. *9*, 283–302.
17. Smyk, M., Obersztyn, E., Nowakowska, B., Bocian, E., Cheung, S.W., Mazurczak, T., and Stankiewicz, P. (2007). Recurrent *SOX9* deletion campomelic dysplasia due to somatic mosaicism in the father. Am. J. Med. Genet. A. *143A*, 866–870.
18. Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D., and Lupski, J.R. (2009). The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat. Genet. *41*, 849–853.
19. Campbell, I.M., Kolodziejska, K.E., Quach, M.M., Wolf, V.L., Cheung, S.-W., Lalani, S.R., Ramocki, M.B., and Stankiewicz, P. (2011). *TGFBR2* deletion in a 20-month-old female with developmental delay and microcephaly. Am. J. Med. Genet. A. *155A*, 1442–1447.
20. Dolezel, J., Bartos, J., Voglmayr, H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. Cytometry A *51*, 127–128, author reply 129.
21. Olofsson, P., and Shaw, C.A. (2002). Exact sampling formulas for multi-type Galton-Watson processes. J. Math. Biol. *45*, 279–293.
22. Drost, J.B., and Lee, W.R. (1995). Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. Environ. Mol. Mutagen. *25* (*Suppl 26*), 48–64.
23. Qin, J., Calabrese, P., Tiemann-Boege, I., Shinde, D.N., Yoon, S.-R., Gelfand, D., Bauer, K., and Arnheim, N. (2007). The molecular anatomy of spontaneous germline mutations in human testes. PLoS Biol. *5*, e224.
24. Joffe, A., and Waugh, W.A.O. (1982). Exact Distributions of Kin Numbers in a Galton-Watson Process. J. Appl. Probab. *19*, 767–775.
25. Kang, S.-H.L., Shaw, C., Ou, Z., Eng, P.A., Cooper, M.L., Pursley, A.N., Sahoo, T., Bacino, C.A., Chinault, A.C., Stankiewicz, P., et al. (2010). Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results. Am. J. Med. Genet. A. *152A*, 1111–1126.
26. Fujimoto, T., Miyayama, Y., and Fuyuta, M. (1977). The origin, migration and fine morphology of human primordial germ cells. Anat. Rec. *188*, 315–330.
27. Marques-Mari, A.I., Lacham-Kaplan, O., Medrano, J.V., Pellicer, A., and Simón, C. (2009). Differentiation of germ cells and gametes from stem cells. Hum. Reprod. Update *15*, 379–390.

28. Crow, J.F. (2000). The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. *1*, 40–47.
29. Röthlisberger, B., and Kotzot, D. (2007). Recurrence risk in *de novo* structural chromosomal rearrangements. Am. J. Med. Genet. A. *143A*, 1708–1714.
30. Hehir-Kwa, J.Y., Rodríguez-Santiago, B., Vissers, L.E., de Leeuw, N., Pfundt, R., Buitelaar, J.K., Pérez-Jurado, L.A., and Veltman, J.A. (2011). *De novo* copy number variants associated with intellectual disability have a paternal origin and age bias. J. Med. Genet. *48*, 776–778.
31. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of *de novo* mutations and the importance of father's age to disease risk. Nature *488*, 471–475.
32. Helderman-van den Enden, A.T.J.M., de Jong, R., den Dunnen, J.T., Houwing-Duistermaat, J.J., Kneppers, A.L., Ginjaar, H.B., Breuning, M.H., and Bakker, E. (2009). Recurrence risk due to germ line mosaicism: Duchenne and Becker muscular dystrophy. Clin. Genet. *75*, 465–472.
33. Sundberg, K., Bang, J., Smidt-Jensen, S., Brocks, V., Lundsteen, C., Parner, J., Keiding, N., and Philip, J. (1997). Randomised study of risk of fetal loss related to early amniocentesis versus chorionic villus sampling. Lancet *350*, 697–703.